RECEIVED
CENTRAL FAX CENTER
MAY 2 1 2004

OFFICIAL

Gates & Cooper LLP
Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045

EXPEDITED PROCEDURE

FAX TRANSMISSION TO USPTO

TO: Commissioner for Patents
Attn: Examiner Cindy Nguyen
Patent Examining Corps
Facsimile Center
Alexandria, VA 22313-1450

FROM: George H. Gates
OUR REF.: STL919990184US3
TELEPHONE: (310) 642-4146

Total pages, including cover letter: **51**

PTO FAX NUMBER: (703) 872-9306

If you do NOT receive all of the pages, please telephone us at (310) 641-8797, or fax us at (310) 641-8798.

Title of Document Transmitted:	**BRIEF OF APPELLANT IN TRIPLICATE AND AUTHORIZATION TO CHARGE DEPOSIT ACCOUNT NO. 09-0460 IN THE AMOUNT OF $330.00 FOR FILING FEE.**
Applicant:	David E. Simmen
Serial No.:	09/669,556
Filed:	September 26, 2000
Group Art Unit:	2171
Our Ref. No.:	STL919990184US3

Please charge all fees to Deposit Account No. 09-0460 of IBM Corporation, the assignee of the present application.

By: [signature]
Name: George H. Gates
Reg. No.: 33,500

I hereby certify that this paper is being transmitted by facsimile to the U.S. Patent and Trademark Office on the date shown below.

[signature]
Signature

5/21/04
Date

GHG/sjm

G&C 30571.272-US-U1

Confirmati n No.: 4709
Due Date: May 23, 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	David E. Simmen	Examiner:	Cindy Nguyen
Serial No.:	09/669,556	Group Art Unit:	2171
Filed:	September 26, 2000	Docket:	STL919990184US3
Title:	QUERY OPTIMIZATION TECHNIQUE FOR OBTAINING IMPROVED CARDINALITY ESTIMATES USING STATISTICS ON PRE-DEFINED QUERIES		

CERTIFICATE OF MAILING OR TRANSMISSION UNDER 37 CFR 1.8

I hereby certify that this correspondence is being filed via facsimile transmission to the U.S. Patent and Trademark Office on May 21, 2004.

By: [signature]
Name: George H. Gates

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

We are transmitting herewith the attached:

☒ Transmittal sheet, in duplicate, containing a Certificate of Mailing or Transmission under 37 CFR 1.8.
☒ BRIEF OF APPELLANT in triplicate.
☒ Charge Deposit Account No. 09-0460 in the amount of $330.00 for Brief of Appellant fee.

Please consider this a PETITION FOR EXTENSION OF TIME for a sufficient number of months to enter these papers, if appropriate.

Please charge all fees to Deposit Account No. 09-0460 of IBM Corporation, the assignee of the present application. A duplicate of this paper is enclosed.

CUSTOMER NUMBER 22462
GATES & COOPER LLP
Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, CA 90045
(310) 641-8797

By: [signature]
Name: George H. Gates
Reg. No.: 33,500
GHG/sjm

Confirmation No.: 4709
Due Date: May 23, 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	David E. Simmen	Examiner:	Cindy Nguyen
Serial No.:	09/669,556	Group Art Unit:	2171
Filed:	September 26, 2000	Docket:	STL919990184US3
Title:	QUERY OPTIMIZATION TECHNIQUE FOR OBTAINING IMPROVED CARDINALITY ESTIMATES USING STATISTICS ON PRE-DEFINED QUERIES		

CERTIFICATE OF MAILING OR TRANSMISSION UNDER 37 CFR 1.8

I hereby certify that this correspondence is being filed via facsimile transmission to the U.S. Patent and Trademark Office on May 21, 2004.

By: [signature]
Name: George H. Gates

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

We are transmitting herewith the attached:

☒ Transmittal sheet, in duplicate, containing a Certificate of Mailing or Transmission under 37 CFR 1.8.
☒ BRIEF OF APPELLANT in triplicate.
☒ Charge Deposit Account No. 09-0460 in the amount of $330.00 for Brief of Appellant fee.

Please consider this a PETITION FOR EXTENSION OF TIME for a sufficient number of months to enter these papers, if appropriate.

Please charge all fees to Deposit Account No. 09-0460 of IBM Corporation, the assignee of the present application. A duplicate of this paper is enclosed.

CUSTOMER NUMBER 22462
GATES & COOPER LLP
Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, CA 90045
(310) 641-8797

By: [signature]
Name: George H. Gates
Reg. No.: 33,500
GHG/sjm

G&C 30571.272-US-U1

Confirmation No.: 4709
Due Date: May 23, 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of:

Inventor: David E. Simmen

Serial #: 09/669,556

Filed: September 26, 2000

Title: QUERY OPTIMIZATION TECHNIQUE FOR OBTAINING IMPROVED CARDINALITY ESTIMATES USING STATISTICS ON PRE-DEFINED QUERIES

Examiner: Cindy Nguyen

Group Art Unit: 2171

Appeal No.: ________

BRIEF OF APPELLANT

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In accordance with 37 C.F.R. §1.192, Appellant's attorney hereby submits the Brief of Appellant, in triplicate, on appeal from the final rejection in the above-identified application as set forth in the Office Action dated December 24, 2003.

Please charge the amount of $330.00 to cover the required fee for filing this Appeal Brief as set forth under 37 C.F.R. §1.17(c) to Deposit Account No. 09-0460 of IBM Corporation the assignee of the present application. Also, please charge any additional fees or credit any overpayments to Deposit Account No. 09-0460.

I. REAL PARTY IN INTEREST

The real party in interest is IBM Corporation, the assignee of the present application.

II. RELATED APPEALS AND INTERFERENCES

There are no related appeals or interferences for the above-referenced patent application.

III. STATUS OF CLAIMS

Claims 1, 3-11, 13-21, and 23-30 are pending in the application. Claims 2, 12, and 22 have been canceled.

Claims 1, 3-7, 11, 13-17, 21, and 23-27 were rejected under 35 U.S.C. §103(a) as being obvious over Schiefer, U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang).

Claims 6-10, 16-20, and 26-30 were rejected under 35 U.S.C. §103(a) as being unpatentable over Schiefer in view of Chiang and further in view of Riatto et al., U.S. Patent No. 5,991,754 (Riatto).

IV. STATUS OF AMENDMENTS

No amendments to the claims have been made subsequent to the final Office Action.

V. SUMMARY OF THE INVENTION

Appellant's invention, as recited in independent claims 1, 11, and 21, is generally directed to a method of optimizing execution of a query that accesses data stored on a data store connected to a computer. Claim 1 is representative and recites the steps of generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and using the generated cardinality estimates to determine an optimal query execution plan for the query.

With regard to the rejected claims, refer to the specification as follows:

(a) at page 6, line 18 through page 29, line 24; and

(b) at page 30, line 1 through page 31, line 17 and in FIGS. 2, 3 and 4 as reference numbers 200-204, 300-310 and 400-416.

VI. ISSUES PRESENTED FOR REVIEW

1. Whether claims 1, 3-7, 11, 13-17, 21, and 23-27 are obvious under 35 U.S.C. §103(a)

in view of Schiefer, U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang).

2. Whether claims 6-10, 16-20, and 26-30 are obvious under 35 U.S.C. §103(a) in view of Schiefer, in view of Chiang and further in view of Riatto et al., U.S. Patent No. 5,991,754 (Riatto).

VII. GROUPING OF CLAIMS

The rejected claims do not all stand or fall together. The claims are grouped as follows:

1. claims 1, 7, 11, 17, 21 and 27 stand or fall together;
2. claims 3, 13 and 23 stand or fall together;
3. claims 4, 14 and 24 stand or fall together;
4. claims 5, 15 and 25 stand or fall together;
5. claims 6, 16 and 26 stand or fall together;
6. claims 8, 18 and 28 stand or fall together;
7. claims 9, 19 and 29 stand or fall together; and
8. claims 10, 20 and 30 stand or fall together.

Separate arguments for each of the groups of claims are provided below.

VIII. ARGUMENTS

A. The Office Action Rejections

In sections (2)-(3) of the Office Action, claims 1, 3-7, 11, 13-17, 21 and 23-27 were rejected under 35 U.S.C. §103(a) as being obvious over Schiefer et al., U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang). In section (4) of the Office Action, claims 6-10, 16-20, and 26-30 were rejected under 35 U.S.C. §103(a) as being unpatentable over Schiefer in view of Chiang, U.S. Patent No. 6,477,523 (Chiang) and further in view of Raitto et al., U.S. Patent No. 5,991,754 (Raitto).

Appellant's attorney respectfully traverses these rejections.

B. The Appellant's Claimed Invention

Appellant's claimed invention, as recited in independent claims 1, 11, and 21, is generally directed to a method of optimizing execution of a query that accesses data stored on a data store

conn cted to a computer. Claim 1 is representative and recites th steps of generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and using the generated cardinality estimates to determine an optimal query execution plan for the query.

C. The Schiefer Reference

Schiefer describes a method for estimating cardinalities for query processing in a relational database management system. The method is suitable for use with a query optimizer for estimating cardinalities for sets of columns or keys resulting from a grouping operation or a duplicate removal operation.

D. The Chiang Reference

Chiang describes a method, apparatus, and article of manufacture for generating statistics for use by a relational database management system. A global aggregate spool is generated for each of a plurality of partitions of a subject table that are spread across a plurality of processing units of a computer system. Each of the global aggregate spools is scanned to generate summary records. The summary records are then merged to generate interval records for a compressed histogram of the subject table, wherein the compressed histogram includes both equal-height intervals and high-biased intervals. The compressed histogram can then be analyzed to estimate the cardinality associated with one or more search conditions of a user query or other SQL statement. Compared to a strictly equal-height histogram, the compressed histogram allows the relational database management system to more accurately estimate the cardinality associated with various search conditions. As a result, the relational database management system can better optimize the execution of the user query.

E. The Raitto Reference

Raitto describes a method and system for processing queries, where the queries do not reference a particular materialized view. Specifically, techniques are provided for handling a query that specifies a first set of one or more aggregate functions, where the particular materialized view reflects a second set of one or more aggregate functions. Whether the query can be rewritten is

determined based on the aggregate functions in the first and second sets, and the corresponding arguments. Techniques are also provided for processing a query that (1) does not reference a particular materialized view, (2) specifies a first set of one or more aggregate functions, where the particular materialized view reflects a second set of one or more aggregate functions. A technique is also provided for rewriting queries that specify an outer join that has a dimension table on the child-side of the outer join and a fact table on the parent-side of the outer join. The query is rewritten to produce a rewritten query by replacing references to the fact table in the query with references to a materialized view. The rewritten query specifies an outer join that has the dimension table on the child side and the materialized view on the parent side.

F. Appellant's Independent Claims Are Patentable Over The Cited References

Appellant's independent claims are patentable over Schiefer, Chiang and Raitto, because it includes a combination of limitations not taught or suggested by the cited references, taken individually or in any combination.

The combination of Schiefer and Chiang is cited by the Office Action as teaching all of the steps or elements of the independent claims 1, 11 and 21.

Appellant's attorney disagrees.

The Office Action states that Chiang teaches the elements "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query" at col. 6, lines 32-65 and in FIG. 3, steps 300-310. However, at the indicated locations, Chiang merely states the following:

Chiang: Col. 6, lines 32-65

According to the preferred embodiment of the present invention, a new kind of database statistics, known as a compressed histogram, are generated for use by the Optimizer subsystem of the PE 114 in optimizing an execution plan. The compressed histogram includes high-biased intervals and/or equal-height intervals that allow the Optimizer subsystem of the PE 114 to more accurately estimate the cardinality associated with various conditions of the execution plan.

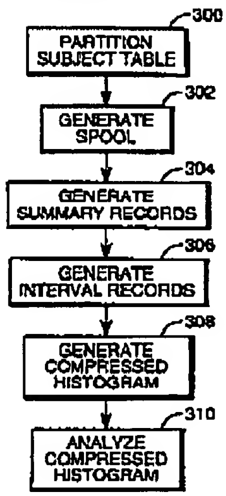
Typically, the compressed histogram is independently generated for a specified subject table and then stored as a single field of a row in a system table in the relational database 118 for later use by the Optimizer subsystem of the PE 114. The PE 114 is responsible for generating the compressed histogram, using a sequence of collection steps sent to and performed by the AMPs 116. In the preferred embodiment, there are two statistics collection steps.

A first collection step is responsible for building a global aggregate spool and a sequence of summary records on each AMP 116 participating in the statistics collection (i.e., on each AMP 116 that manages a partition of the subject table), wherein multiple copies of the first collection step are executed simultaneously and in parallel by the AMPs 116. In this manner, the global aggregate spool may be considered partitioned in the same manner as the subject table.

Each row of the global aggregate spool includes: (1) a distinct value from the partition of the subject table and (2) the number of rows in the partition of the subject table having the distinct value. The global aggregate spool is considered global in the sense that a distinct value from the subject table can only be found on a single AMP 116, because the subject table is partitioned across multiple AMPs 116.

Chiang: FIG. 3

FIG. 3



Nothing in the above description from Chiang can fairly be said to represent "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query."

In Chiang, summary records are constructed from a global aggregate spool. Each row of the global aggregate spool includes: (1) a distinct value from the partition of the subject table and (2) the number of rows in the partition of the subject table having the distinct value. Each summary record includes: (1) a sort key, (2) a distinct value, and (3) the number of rows in the partition of the subject table having the distinct value.

However, the summary records in Chiang are not "automatic summary tables" or "materialized views." As noted in Appellant's specification, automatic summary tables are pre-computed queries.

Also, Chiang does not determine that an automatic summary table vertically overlaps a query. As noted in Appellant's specification, an automatic summary table vertically overlaps a query when the set of predicates applied by the automatic summary table is a subset of the predicates required by the query.

However, there is no discussion of vertically overlapping automatic summary tables in Chiang. Indeed, Chiang is directed only to the construction of a compressed histogram of a subject table without reference to a query.

Consequently, Chiang does not teach or suggest "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query."

The Office Action also states that Schiefer teaches the elements "using the cardinality estimates to determine an optimal query execution plan for the query" at col. 3, lines 37-60 and col. 10, lines 23-57. However, at the indicated locations, Schiefer merely states the following:

> Schiefer: Col. 3, lines 37-60
>
> It is another object of the present invention to produce a better cardinality estimate by utilizing information and attributes which can be obtained from the catalog for the relational database management system. The additional information includes cardinalities for existing unique keys, column equivalence classes, functional dependencies, statistical functional dependencies, and statistically unique keys.
>
> In a first aspect, the present invention provides a method for estimating cardinalities for a key formed from a grouping of columns in a table for use in a query optimizer for a relational database management system, wherein selectivities and keys associated with columns in the table are provided in a catalog, said method comprising the steps of: (a) determining an equivalence class for each column in said key; (b) for each said equivalence class determining an effective cardinality for each of said columns belonging to said equivalence class; (c) determining a cardinality for each of said equivalence classes by choosing the minimum effective cardinality for the columns belonging to said equivalence class; and (d) estimating a cardinality value for said key from the product of said cardinalities for said equivalence classes.
>
> Schiefer: Col. 10, lines 23-57
>
> To determine the effective cardinality of a column in Line 13, the method according to the present invention considers the effect of local predicates on other columns in the equivalence class. Known query optimizers estimate the cardinality of a column C1 using only the product of predicate selectivity (ff_1) and base table column cardinality .vertline.C1.vertline. obtained from the CATALOG. Known optimizers do not consider the effects of predicates on other columns. According to the invention, the effective cardinality of a column is determined by the following expression which will be referred to as Expression (1):

EFFECTIVE COLUMN CARDINALITY = $|C1| * ff\_1 * (1 - (1 - ff\_2)^{(|T|/|C1|)})$ (1)

where:

$|T|$ is the table cardinality, i.e. number of rows in table

$|C1|$ is the base table cardinality obtained from the CATALOG

ff_1 is the selectivity of a local predicate for column C1

ff_2 is the selectively of a local predicate for column C2

In the derivation of Expression (1) according to the present method, it is assumed that C1 and C2 are independent, and the values of C1 and C2 are both uniformly distributed in the table.

If there is no restriction on column C2, i.e. ff_2 is 1, Expression (1) reduces to $|C1| * ff\_1$ which provides the basic operation performed by known optimizers for obtaining the effective cardinality of a column. Since the prior art method is based on the assumption that all columns in a key are fully independent of each other, the method according to the prior art usually leads to unnecessarily large numbers for the key cardinalities. This in turn can result in the query optimizer 18 (FIG. 1) picking the wrong query plan which is clearly undesirable.

In the context of Appellant's claims, the cardinality estimates are generated using statistics of one or more automatic summary tables that vertically overlap the query. In Schiefer, however, the cardinality estimates are generated by (1) determining an equivalence class for each column in a key; (b) for each equivalence class, determining an effective cardinality for each of the columns belonging to the equivalence class; (c) determining a cardinality for each of the equivalence classes by choosing the minimum effective cardinality for the columns belonging to the equivalence class; and (d) estimating a cardinality value for the key from the product of the cardinalities for the equivalence classes.

Raitto does not overcome the deficiencies of Schiefer and Chiang. Recall that Raitto was cited only against dependent claims 6-10, 16-20 and 26-30, and is specifically directed to queries that do not reference a particular materialized view (automatic summary table).

Consequently, even when combined, the Schiefer, Chiang and Raitto references do not teach or suggest the Appellant's invention. Moreover, the various elements of Appellant's claimed invention together provide operational advantages over the cited references. In addition, Appellant's invention solves problems not recognized by the cited references.

Thus, Appellant submits that independent claims 1, 11 and 21 are allowable over Schiefer, Chiang and Raitto. Further, dependent claims 3-10, 13-20 and 23-30 are submitted to be allowable

over Schiefer, Chiang and Raitto in the same manner, because they are dependent on independent claims 1, 11 and 21, respectively, and because they contain all the limitations of the independent claims.

G. Appellant's Dependent Claims Are Patentable Over The Cited References

Appellant's dependent claims are patentable over Schiefer, Chiang and Raitto, because it includes a combination of limitations not taught or suggested by the cited references, taken individually or in any combination

With regard to claims 3, 13 and 23, which recite that "the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query," the Office Action asserts that Schiefer teaches these elements at col. 6, line 41 – col. 7, line 20. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes estimating cardinalities, but not using statistics of automatic summary tables.

With regard to claims 4, 14 and 24, which recite that "the predicates are applied by one of the automatic summary tables," the Office Action asserts that Schiefer teaches these elements at col. 10, lines 23-44. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes the effect of local predicates on other columns, but not the application of predicates by automatic summary tables.

With regard to claims 5, 15 and 25, which recite that "the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query," the Office Action asserts that Schiefer teaches these elements at col. 8, lines 1-28. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes estimates of key cardinalities, but says nothing about a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

With regard to claims 6, 16 and 24, which recite that "zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table," the Office Action asserts that Chiang teaches these elements at col. 12, lines 28-67. Appellant's attorney disagrees. The indicated location in Chiang is completely unrelated to the claim limitations.

With regard to claims 7, 17 and 27, these claims stand or fall with claims 1, 11 and 21.

With regard to claims 8, 18 and 28, which recite "determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table," the Office Action asserts that Raitto teaches these elements at col. 11, lines 6-19. Appellant's attorney disagrees. The indicated location in Raitto describes a "query reduction factor" computed for a materialized view, which is a different ratio, comprising a ratio of (1) the sum of the cardinalities of matching relations in the query that will be replaced by the materialized view to (2) the cardinality of the materialized view.

With regard to claims 9, 19 and 29, which recite that "a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query," the Office Action asserts that Raitto teaches these elements at col. 11, lines 20-30. Appellant's attorney disagrees. At the indicated location, Raitto describes a "query reduction factor" computed for a materialized view, which is a different ratio, comprising a ratio of (1) the sum of the cardinalities of matching relations in the query that will be replaced by the materialized view to (2) the cardinality of the materialized view.

With regard to claims 10, 20 and 30, which recite that "the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio," the Office Action asserts that Raitto teaches these elements at col. 11, lines 31-49. Appellant's attorney disagrees. At the indicated location, Raitto merely describes the "current best materialized view" with the "highest query reduction factor," but not a product of the subpredicate combined selectivity estimate and the cardinality ratio.

IX. CONCLUSION

In light of the above arguments, Appellant respectfully submits that the cited references do not anticipate nor render obvious the claimed invention. More specifically, Appellant's claims recite novel physical features, which patentably distinguish over any and all references under 35 U.S.C. §§ 102 and 103. As a result, a decision by the Board of Patent Appeals and Interferences reversing the Examiner and directing allowance of the pending claims in the subject application is respectfully solicited.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Appellant

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: May 21, 2004

By: ____________________
Name: George H. Gates
Reg. No.: 33,500

GHG/

G&C 30571.272-US-U1

APPENDIX

1. A method of optimizing execution of a query that accesses data stored on a data store connected to a computer, comprising:

generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query; and

using the generated cardinality estimates to determine an optimal query execution plan for the query.

3. The method of claim 1, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

4. The method of claim 3, wherein the predicates are applied by one of the automatic summary tables.

5. The method of claim 4, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

6. The method of claim 3, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

7. The method of claim 6, wherein a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

8. The method of claim 7, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

9. The method of claim 8, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

10. The method of claim 9, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

11. An apparatus for optimizing execution of a query, comprising:

a computer having a data store coupled thereto, wherein the data store stores data;

one or more computer programs, performed by the computer, for generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and for using the generated cardinality estimates to determine an optimal query execution plan for the query.

13. The apparatus of claim 11, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

14. The apparatus of claim 13, wherein the predicates are applied by one of the automatic summary tables.

15. The apparatus of claim 14, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

16. The apparatus of claim 13, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

17. The apparatus of claim 16, a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

18. The apparatus of claim 17, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

19. The apparatus of claim 18, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

20. The apparatus of claim 19, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

21. An article of manufacture comprising a program storage medium readable by a computer and embodying one or more instructions executable by the computer to optimizing execution of a query that accesses data stored on a data store connected to the computer, comprising:

generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query; and

using the generated cardinality estimates to determine an optimal query execution plan for the query.

23. The article of manufacture of claim 21, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

24. The article of manufacture of claim 23, wherein the predicates are applied by one of the automatic summary tables.

25. The article of manufacture of claim 24, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

26. The article of manufacture of claim 23, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

27. The article of manufacture of claim 26, a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

28. The article of manufacture of claim 27, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

29. The article of manufacture of claim 28, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

30. The article of manufacture of claim 29, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

Confirmation No.: 4709
Due Date: May 23, 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of:	
Inventor: David E. Simmen	Examiner: Cindy Nguyen
Serial #: 09/669,556	Group Art Unit: 2171
Filed: September 26, 2000	Appeal No.: ________
Title: QUERY OPTIMIZATION TECHNIQUE FOR OBTAINING IMPROVED CARDINALITY ESTIMATES USING STATISTICS ON PRE-DEFINED QUERIES	

BRIEF OF APPELLANT

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In accordance with 37 C.F.R. §1.192, Appellant's attorney hereby submits the Brief of Appellant, in triplicate, on appeal from the final rejection in the above-identified application as set forth in the Office Action dated December 24, 2003.

Please charge the amount of $330.00 to cover the required fee for filing this Appeal Brief as set forth under 37 C.F.R. §1.17(c) to Deposit Account No. 09-0460 of IBM Corporation the assignee of the present application. Also, please charge any additional fees or credit any overpayments to Deposit Account No. 09-0460.

I. REAL PARTY IN INTEREST

The real party in interest is IBM Corporation, the assignee of the present application.

II. RELATED APPEALS AND INTERFERENCES

There are no related appeals or interferences for the above-referenced patent application.

III. STATUS OF CLAIMS

Claims 1, 3-11, 13-21, and 23-30 are pending in the application. Claims 2, 12, and 22 have been canceled.

Claims 1, 3-7, 11, 13-17, 21, and 23-27 were rejected under 35 U.S.C. §103(a) as being obvious over Schiefer, U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang).

Claims 6-10, 16-20, and 26-30 were rejected under 35 U.S.C. §103(a) as being unpatentable over Schiefer in view of Chiang and further in view of Riatto et al., U.S. Patent No. 5,991,754 (Riatto).

IV. STATUS OF AMENDMENTS

No amendments to the claims have been made subsequent to the final Office Action.

V. SUMMARY OF THE INVENTION

Appellant's invention, as recited in independent claims 1, 11, and 21, is generally directed to a method of optimizing execution of a query that accesses data stored on a data store connected to a computer. Claim 1 is representative and recites the steps of generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and using the generated cardinality estimates to determine an optimal query execution plan for the query.

With regard to the rejected claims, refer to the specification as follows:

(a) at page 6, line 18 through page 29, line 24; and

(b) at page 30, line 1 through page 31, line 17 and in FIGS. 2, 3 and 4 as reference numbers 200-204, 300-310 and 400-416.

VI. ISSUES PRESENTED FOR REVIEW

1. Whether claims 1, 3-7, 11, 13-17, 21, and 23-27 are obvious under 35 U.S.C. §103(a)

in view of Schiefer, U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang).

2. Whether claims 6-10, 16-20, and 26-30 are obvious under 35 U.S.C. §103(a) in view of Schiefer, in view of Chiang and further in view of Riatto et al., U.S. Patent No. 5,991,754 (Riatto).

VII. GROUPING OF CLAIMS

The rejected claims do not all stand or fall together. The claims are grouped as follows:

1. claims 1, 7, 11, 17, 21 and 27 stand or fall together;
2. claims 3, 13 and 23 stand or fall together;
3. claims 4, 14 and 24 stand or fall together;
4. claims 5, 15 and 25 stand or fall together;
5. claims 6, 16 and 26 stand or fall together;
6. claims 8, 18 and 28 stand or fall together;
7. claims 9, 19 and 29 stand or fall together; and
8. claims 10, 20 and 30 stand or fall together.

Separate arguments for each of the groups of claims are provided below.

VIII. ARGUMENTS

A. The Office Action Rejections

In sections (2)-(3) of the Office Action, claims 1, 3-7, 11, 13-17, 21 and 23-27 were rejected under 35 U.S.C. §103(a) as being obvious over Schiefer et al., U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang). In section (4) of the Office Action, claims 6-10, 16-20, and 26-30 were rejected under 35 U.S.C. §103(a) as being unpatentable over Schiefer in view of Chiang, U.S. Patent No. 6,477,523 (Chiang) and further in view of Raitto et al., U.S. Patent No. 5,991,754 (Raitto).

Appellant's attorney respectfully traverses these rejections.

B. The Appellant's Claimed Invention

Appellant's claimed invention, as recited in independent claims 1, 11, and 21, is generally directed to a method of optimizing execution of a query that accesses data stored on a data store

connected to a computer. Claim 1 is representative and recites the steps of generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and using the generated cardinality estimates to determine an optimal query execution plan for the query.

C. The Schiefer Reference

Schiefer describes a method for estimating cardinalities for query processing in a relational database management system. The method is suitable for use with a query optimizer for estimating cardinalities for sets of columns or keys resulting from a grouping operation or a duplicate removal operation.

D. The Chiang Reference

Chiang describes a method, apparatus, and article of manufacture for generating statistics for use by a relational database management system. A global aggregate spool is generated for each of a plurality of partitions of a subject table that are spread across a plurality of processing units of a computer system. Each of the global aggregate spools is scanned to generate summary records. The summary records are then merged to generate interval records for a compressed histogram of the subject table, wherein the compressed histogram includes both equal-height intervals and high-biased intervals. The compressed histogram can then be analyzed to estimate the cardinality associated with one or more search conditions of a user query or other SQL statement. Compared to a strictly equal-height histogram, the compressed histogram allows the relational database management system to more accurately estimate the cardinality associated with various search conditions. As a result, the relational database management system can better optimize the execution of the user query.

E. The Raitto Reference

Raitto describes a method and system for processing queries, where the queries do not reference a particular materialized view. Specifically, techniques are provided for handling a query that specifies a first set of one or more aggregate functions, where the particular materialized view reflects a second set of one or more aggregate functions. Whether the query can be rewritten is

determined based on the aggregate functions in the first and second sets, and the corresponding arguments. Techniques are also provided for processing a query that (1) does not reference a particular materialized view, (2) specifies a first set of one or more aggregate functions, where the particular materialized view reflects a second set of one or more aggregate functions. A technique is also provided for rewriting queries that specify an outer join that has a dimension table on the child-side of the outer join and a fact table on the parent-side of the outer join. The query is rewritten to produce a rewritten query by replacing references to the fact table in the query with references to a materialized view. The rewritten query specifies an outer join that has the dimension table on the child side and the materialized view on the parent side.

F. Appellant's Independent Claims Are Patentable Over The Cited References

Appellant's independent claims are patentable over Schiefer, Chiang and Raitto, because it includes a combination of limitations not taught or suggested by the cited references, taken individually or in any combination.

The combination of Schiefer and Chiang is cited by the Office Action as teaching all of the steps or elements of the independent claims 1, 11 and 21.

Appellant's attorney disagrees.

The Office Action states that Chiang teaches the elements "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query" at col. 6, lines 32-65 and in FIG. 3, steps 300-310. However, at the indicated locations, Chiang merely states the following:

Chiang: Col. 6, lines 32-65

According to the preferred embodiment of the present invention, a new kind of database statistics, known as a compressed histogram, are generated for use by the Optimizer subsystem of the PE 114 in optimizing an execution plan. The compressed histogram includes high-biased intervals and/or equal-height intervals that allow the Optimizer subsystem of the PE 114 to more accurately estimate the cardinality associated with various conditions of the execution plan.

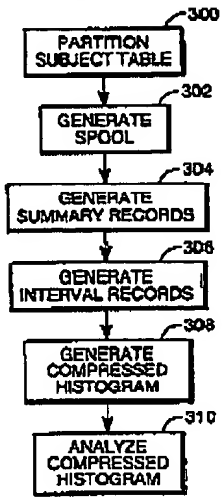
Typically, the compressed histogram is independently generated for a specified subject table and then stored as a single field of a row in a system table in the relational database 118 for later use by the Optimizer subsystem of the PE 114. The PE 114 is responsible for generating the compressed histogram, using a sequence of collection steps sent to and performed by the AMPs 116. In the preferred embodiment, there are two statistics collection steps.

A first collection step is responsible for building a global aggregate spool and a sequence of summary records on each AMP 116 participating in the statistics collection (i.e., on each AMP 116 that manages a partition of the subject table), wherein multiple copies of the first collection step are executed simultaneously and in parallel by the AMPs 116. In this manner, the global aggregate spool may be considered partitioned in the same manner as the subject table.

Each row of the global aggregate spool includes: (1) a distinct value from the partition of the subject table and (2) the number of rows in the partition of the subject table having the distinct value. The global aggregate spool is considered global in the sense that a distinct value from the subject table can only be found on a single AMP 116, because the subject table is partitioned across multiple AMPs 116.

Chiang: FIG. 3

FIG. 3



Nothing in the above description from Chiang can fairly be said to represent "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query."

In Chiang, summary records are constructed from a global aggregate spool. Each row of the global aggregate spool includes: (1) a distinct value from the partition of the subject table and (2) the number of rows in the partition of the subject table having the distinct value. Each summary record includes: (1) a sort key, (2) a distinct value, and (3) the number of rows in the partition of the subject table having the distinct value.

However, the summary records in Chiang are not "automatic summary tables" or "materialized views." As noted in Appellant's specification, automatic summary tables are pre-computed queries.

Also, Chiang does not determine that an automatic summary table vertically overlaps a query. As noted in Appellant's specification, an automatic summary table vertically overlaps a query when the set of predicates applied by the automatic summary table is a subset of the predicates required by the query.

However, there is no discussion of vertically overlapping automatic summary tables in Chiang. Indeed, Chiang is directed only to the construction of a compressed histogram of a subject table without reference to a query.

Consequently, Chiang does not teach or suggest "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query."

The Office Action also states that Schiefer teaches the elements "using the cardinality estimates to determine an optimal query execution plan for the query" at col. 3, lines 37-60 and col. 10, lines 23-57. However, at the indicated locations, Schiefer merely states the following:

> <u>Schiefer: Col. 3, lines 37-60</u>
>
> It is another object of the present invention to produce a better cardinality estimate by utilizing information and attributes which can be obtained from the catalog for the relational database management system. The additional information includes cardinalities for existing unique keys, column equivalence classes, functional dependencies, statistical functional dependencies, and statistically unique keys.
>
> In a first aspect, the present invention provides a method for estimating cardinalities for a key formed from a grouping of columns in a table for use in a query optimizer for a relational database management system, wherein selectivities and keys associated with columns in the table are provided in a catalog, said method comprising the steps of: (a) determining an equivalence class for each column in said key; (b) for each said equivalence class determining an effective cardinality for each of said columns belonging to said equivalence class; (c) determining a cardinality for each of said equivalence classes by choosing the minimum effective cardinality for the columns belonging to said equivalence class; and (d) estimating a cardinality value for said key from the product of said cardinalities for said equivalence classes.
>
> <u>Schiefer: Col. 10, lines 23-57</u>
>
> To determine the effective cardinality of a column in Line 13, the method according to the present invention considers the effect of local predicates on other columns in the equivalence class. Known query optimizers estimate the cardinality of a column C1 using only the product of predicate selectivity (ff_1) and base table column cardinality .vertline.C1.vertline. obtained from the CATALOG. Known optimizers do not consider the effects of predicates on other columns. According to the invention, the effective cardinality of a column is determined by the following expression which will be referred to as Expression (1):

EFFECTIVE COLUMN CARDINALITY = $|C1| * ff\_1 * (1 - (1 - ff\_2)^{(|T|/|C1|)})$(1)

where:

$|T|$ is the table cardinality, i.e. number of rows in table

$|C1|$ is the base table cardinality obtained from the CATALOG

ff_1 is the selectivity of a local predicate for column C1

ff_2 is the selectively of a local predicate for column C2

In the derivation of Expression (1) according to the present method, it is assumed that C1 and C2 are independent, and the values of C1 and C2 are both uniformly distributed in the table.

If there is no restriction on column C2, i.e. ff_2 is 1, Expression (1) reduces to $|C1| * ff\_1$ which provides the basic operation performed by known optimizers for obtaining the effective cardinality of a column. Since the prior art method is based on the assumption that all columns in a key are fully independent of each other, the method according to the prior art usually leads to unnecessarily large numbers for the key cardinalities. This in turn can result in the query optimizer 18 (FIG. 1) picking the wrong query plan which is clearly undesirable.

In the context of Appellant's claims, the cardinality estimates are generated using statistics of one or more automatic summary tables that vertically overlap the query. In Schiefer, however, the cardinality estimates are generated by (1) determining an equivalence class for each column in a key; (b) for each equivalence class, determining an effective cardinality for each of the columns belonging to the equivalence class; (c) determining a cardinality for each of the equivalence classes by choosing the minimum effective cardinality for the columns belonging to the equivalence class; and (d) estimating a cardinality value for the key from the product of the cardinalities for the equivalence classes.

Raitto does not overcome the deficiencies of Schiefer and Chiang. Recall that Raitto was cited only against dependent claims 6-10, 16-20 and 26-30, and is specifically directed to queries that do not reference a particular materialized view (automatic summary table).

Consequently, even when combined, the Schiefer, Chiang and Raitto references do not teach or suggest the Appellant's invention. Moreover, the various elements of Appellant's claimed invention together provide operational advantages over the cited references. In addition, Appellant's invention solves problems not recognized by the cited references.

Thus, Appellant submits that independent claims 1, 11 and 21 are allowable over Schiefer, Chiang and Raitto. Further, dependent claims 3-10, 13-20 and 23-30 are submitted to be allowable

over Schiefer, Chiang and Raitto in the same manner, because they are dependent on independent claims 1, 11 and 21, respectively, and because they contain all the limitations of the independent claims.

G. Appellant's Dependent Claims Are Patentable Over The Cited References

Appellant's dependent claims are patentable over Schiefer, Chiang and Raitto, because it includes a combination of limitations not taught or suggested by the cited references, taken individually or in any combination

With regard to claims 3, 13 and 23, which recite that "the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query," the Office Action asserts that Schiefer teaches these elements at col. 6, line 41 – col. 7, line 20. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes estimating cardinalities, but not using statistics of automatic summary tables.

With regard to claims 4, 14 and 24, which recite that "the predicates are applied by one of the automatic summary tables," the Office Action asserts that Schiefer teaches these elements at col. 10, lines 23-44. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes the effect of local predicates on other columns, but not the application of predicates by automatic summary tables.

With regard to claims 5, 15 and 25, which recite that "the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query," the Office Action asserts that Schiefer teaches these elements at col. 8, lines 1-28. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes estimates of key cardinalities, but says nothing about a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

With regard to claims 6, 16 and 24, which recite that "zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table," the Office Action asserts that Chiang teaches these elements at col. 12, lines 28-67. Appellant's attorney disagrees. The indicated location in Chiang is completely unrelated to the claim limitations.

With regard to claims 7, 17 and 27, these claims stand or fall with claims 1, 11 and 21.

With regard to claims 8, 18 and 28, which recite "determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table," the Office Action asserts that Raitto teaches these elements at col. 11, lines 6-19. Appellant's attorney disagrees. The indicated location in Raitto describes a "query reduction factor" computed for a materialized view, which is a different ratio, comprising a ratio of (1) the sum of the cardinalities of matching relations in the query that will be replaced by the materialized view to (2) the cardinality of the materialized view.

With regard to claims 9, 19 and 29, which recite that "a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query," the Office Action asserts that Raitto teaches these elements at col. 11, lines 20-30. Appellant's attorney disagrees. At the indicated location, Raitto describes a "query reduction factor" computed for a materialized view, which is a different ratio, comprising a ratio of (1) the sum of the cardinalities of matching relations in the query that will be replaced by the materialized view to (2) the cardinality of the materialized view.

With regard to claims 10, 20 and 30, which recite that "the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio," the Office Action asserts that Raitto teaches these elements at col. 11, lines 31-49. Appellant's attorney disagrees. At the indicated location, Raitto merely describes the "current best materialized view" with the "highest query reduction factor," but not a product of the subpredicate combined selectivity estimate and the cardinality ratio.

IX. CONCLUSION

In light of the above arguments, Appellant respectfully submits that the cited references do not anticipate nor render obvious the claimed invention. More specifically, Appellant's claims recite novel physical features, which patentably distinguish over any and all references under 35 U.S.C. §§ 102 and 103. As a result, a decision by the Board of Patent Appeals and Interferences reversing the Examiner and directing allowance of the pending claims in the subject application is respectfully solicited.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Appellant

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: May 21, 2004

By: [signature]
Name: George H. Gates
Reg. No.: 33,500

GHG/

G&C 30571.272-US-U1

APPENDIX

1. A method of optimizing execution of a query that accesses data stored on a data store connected to a computer, comprising:

generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query; and

using the generated cardinality estimates to determine an optimal query execution plan for the query.

3. The method of claim 1, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

4. The method of claim 3, wherein the predicates are applied by one of the automatic summary tables.

5. The method of claim 4, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

6. The method of claim 3, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

7. The method of claim 6, wherein a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

8. The method of claim 7, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

9. The method of claim 8, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

10. The method of claim 9, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

11. An apparatus for optimizing execution of a query, comprising:

a computer having a data store coupled thereto, wherein the data store stores data;

one or more computer programs, performed by the computer, for generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and for using the generated cardinality estimates to determine an optimal query execution plan for the query.

13. The apparatus of claim 11, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

14. The apparatus of claim 13, wherein the predicates are applied by one of the automatic summary tables.

15. The apparatus of claim 14, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

16. The apparatus of claim 13, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

17. The apparatus of claim 16, a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

18. The apparatus of claim 17, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

19. The apparatus of claim 18, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

20. The apparatus of claim 19, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

21. An article of manufacture comprising a program storage medium readable by a computer and embodying one or more instructions executable by the computer to optimizing execution of a query that accesses data stored on a data store connected to the computer, comprising:

generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query; and

using the generated cardinality estimates to determine an optimal query execution plan for the query.

23. The article of manufacture of claim 21, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

24. The article of manufacture of claim 23, wherein the predicates are applied by one of the automatic summary tables.

25. The article of manufacture of claim 24, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

26. The article of manufacture of claim 23, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

27. The article of manufacture of claim 26, a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

28. The article of manufacture of claim 27, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

29. The article of manufacture of claim 28, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

30. The article of manufacture of claim 29, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

Confirmation No.: 4709
Due Date: May 23, 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of:	
Inventor: David E. Simmen	Examiner: Cindy Nguyen
Serial #: 09/669,556	Group Art Unit: 2171
Filed: September 26, 2000	Appeal No.: _________
Title: QUERY OPTIMIZATION TECHNIQUE FOR OBTAINING IMPROVED CARDINALITY ESTIMATES USING STATISTICS ON PRE-DEFINED QUERIES	

BRIEF OF APPELLANT

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In accordance with 37 C.F.R. §1.192, Appellant's attorney hereby submits the Brief of Appellant, in triplicate, on appeal from the final rejection in the above-identified application as set forth in the Office Action dated December 24, 2003.

Please charge the amount of $330.00 to cover the required fee for filing this Appeal Brief as set forth under 37 C.F.R. §1.17(c) to Deposit Account No. 09-0460 of IBM Corporation the assignee of the present application. Also, please charge any additional fees or credit any overpayments to Deposit Account No. 09-0460.

I. REAL PARTY IN INTEREST

The real party in interest is IBM Corporation, the assignee of the present application.

II. RELATED APPEALS AND INTERFERENCES

There are no related appeals or interferences for the above-referenced patent application.

III. STATUS OF CLAIMS

Claims 1, 3-11, 13-21, and 23-30 are pending in the application. Claims 2, 12, and 22 have been canceled.

Claims 1, 3-7, 11, 13-17, 21, and 23-27 were rejected under 35 U.S.C. §103(a) as being obvious over Schiefer, U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang).

Claims 6-10, 16-20, and 26-30 were rejected under 35 U.S.C. §103(a) as being unpatentable over Schiefer in view of Chiang and further in view of Riatto et al., U.S. Patent No. 5,991,754 (Riatto).

IV. STATUS OF AMENDMENTS

No amendments to the claims have been made subsequent to the final Office Action.

V. SUMMARY OF THE INVENTION

Appellant's invention, as recited in independent claims 1, 11, and 21, is generally directed to a method of optimizing execution of a query that accesses data stored on a data store connected to a computer. Claim 1 is representative and recites the steps of generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and using the generated cardinality estimates to determine an optimal query execution plan for the query.

With regard to the rejected claims, refer to the specification as follows:

(a) at page 6, line 18 through page 29, line 24; and

(b) at page 30, line 1 through page 31, line 17 and in FIGS. 2, 3 and 4 as reference numbers 200-204, 300-310 and 400-416.

VI. ISSUES PRESENTED FOR REVIEW

1. Whether claims 1, 3-7, 11, 13-17, 21, and 23-27 are obvious under 35 U.S.C. §103(a)

in view of Schiefer, U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang).

2. Whether claims 6-10, 16-20, and 26-30 are obvious under 35 U.S.C. §103(a) in view of Schiefer, in view of Chiang and further in view of Riatto et al., U.S. Patent No. 5,991,754 (Riatto).

VII. GROUPING OF CLAIMS

The rejected claims do not all stand or fall together. The claims are grouped as follows:

1. claims 1, 7, 11, 17, 21 and 27 stand or fall together;
2. claims 3, 13 and 23 stand or fall together;
3. claims 4, 14 and 24 stand or fall together;
4. claims 5, 15 and 25 stand or fall together;
5. claims 6, 16 and 26 stand or fall together;
6. claims 8, 18 and 28 stand or fall together;
7. claims 9, 19 and 29 stand or fall together; and
8. claims 10, 20 and 30 stand or fall together.

Separate arguments for each of the groups of claims are provided below.

VIII. ARGUMENTS

A. The Office Action Rejections

In sections (2)-(3) of the Office Action, claims 1, 3-7, 11, 13-17, 21 and 23-27 were rejected under 35 U.S.C. §103(a) as being obvious over Schiefer et al., U.S. Patent No. 5,761,653 (Schiefer) in view of Chiang, U.S. Patent No. 6,477,523 (Chiang). In section (4) of the Office Action, claims 6-10, 16-20, and 26-30 were rejected under 35 U.S.C. §103(a) as being unpatentable over Schiefer in view of Chiang, U.S. Patent No. 6,477,523 (Chiang) and further in view of Raitto et al., U.S. Patent No. 5,991,754 (Raitto).

Appellant's attorney respectfully traverses these rejections.

B. The Appellant's Claimed Invention

Appellant's claimed invention, as recited in independent claims 1, 11, and 21, is generally directed to a method of optimizing execution of a query that accesses data stored on a data store

connected to a computer. Claim 1 is representative and recites the steps of generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and using the generated cardinality estimates to determine an optimal query execution plan for the query.

C. The Schiefer Reference

Schiefer describes a method for estimating cardinalities for query processing in a relational database management system. The method is suitable for use with a query optimizer for estimating cardinalities for sets of columns or keys resulting from a grouping operation or a duplicate removal operation.

D. The Chiang Reference

Chiang describes a method, apparatus, and article of manufacture for generating statistics for use by a relational database management system. A global aggregate spool is generated for each of a plurality of partitions of a subject table that are spread across a plurality of processing units of a computer system. Each of the global aggregate spools is scanned to generate summary records. The summary records are then merged to generate interval records for a compressed histogram of the subject table, wherein the compressed histogram includes both equal-height intervals and high-biased intervals. The compressed histogram can then be analyzed to estimate the cardinality associated with one or more search conditions of a user query or other SQL statement. Compared to a strictly equal-height histogram, the compressed histogram allows the relational database management system to more accurately estimate the cardinality associated with various search conditions. As a result, the relational database management system can better optimize the execution of the user query.

E. The Raitto Reference

Raitto describes a method and system for processing queries, where the queries do not reference a particular materialized view. Specifically, techniques are provided for handling a query that specifies a first set of one or more aggregate functions, where the particular materialized view reflects a second set of one or more aggregate functions. Whether the query can be rewritten is

determined based on the aggregate functions in the first and second sets, and the corresponding arguments. Techniques are also provided for processing a query that (1) does not reference a particular materialized view, (2) specifies a first set of one or more aggregate functions, where the particular materialized view reflects a second set of one or more aggregate functions. A technique is also provided for rewriting queries that specify an outer join that has a dimension table on the child-side of the outer join and a fact table on the parent-side of the outer join. The query is rewritten to produce a rewritten query by replacing references to the fact table in the query with references to a materialized view. The rewritten query specifies an outer join that has the dimension table on the child side and the materialized view on the parent side.

F. Appellant's Independent Claims Are Patentable Over The Cited References

Appellant's independent claims are patentable over Schiefer, Chiang and Raitto, because it includes a combination of limitations not taught or suggested by the cited references, taken individually or in any combination.

The combination of Schiefer and Chiang is cited by the Office Action as teaching all of the steps or elements of the independent claims 1, 11 and 21.

Appellant's attorney disagrees.

The Office Action states that Chiang teaches the elements "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query" at col. 6, lines 32-65 and in FIG. 3, steps 300-310. However, at the indicated locations, Chiang merely states the following:

Chiang: Col. 6, lines 32-65

According to the preferred embodiment of the present invention, a new kind of database statistics, known as a compressed histogram, are generated for use by the Optimizer subsystem of the PE 114 in optimizing an execution plan. The compressed histogram includes high-biased intervals and/or equal-height intervals that allow the Optimizer subsystem of the PE 114 to more accurately estimate the cardinality associated with various conditions of the execution plan.

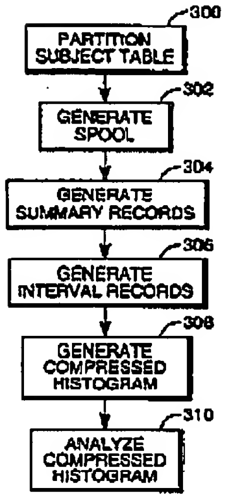
Typically, the compressed histogram is independently generated for a specified subject table and then stored as a single field of a row in a system table in the relational database 118 for later use by the Optimizer subsystem of the PE 114. The PE 114 is responsible for generating the compressed histogram, using a sequence of collection steps sent to and performed by the AMPs 116. In the preferred embodiment, there are two statistics collection steps.

A first collection step is responsible for building a global aggregate spool and a sequence of summary records on each AMP 116 participating in the statistics collection (i.e., on each AMP 116 that manages a partition of the subject table), wherein multiple copies of the first collection step are executed simultaneously and in parallel by the AMPs 116. In this manner, the global aggregate spool may be considered partitioned in the same manner as the subject table.

Each row of the global aggregate spool includes: (1) a distinct value from the partition of the subject table and (2) the number of rows in the partition of the subject table having the distinct value. The global aggregate spool is considered global in the sense that a distinct value from the subject table can only be found on a single AMP 116, because the subject table is partitioned across multiple AMPs 116.

Chiang: FIG. 3

FIG. 3



Nothing in the above description from Chiang can fairly be said to represent "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query."

In Chiang, summary records are constructed from a global aggregate spool. Each row of the global aggregate spool includes: (1) a distinct value from the partition of the subject table and (2) the number of rows in the partition of the subject table having the distinct value. Each summary record includes: (1) a sort key, (2) a distinct value, and (3) the number of rows in the partition of the subject table having the distinct value.

However, the summary records in Chiang are not "automatic summary tables" or "materialized views." As noted in Appellant's specification, automatic summary tables are pre-computed queries.

Also, Chiang does not determine that an automatic summary table vertically overlaps a query. As noted in Appellant's specification, an automatic summary table vertically overlaps a query when the set of predicates applied by the automatic summary table is a subset of the predicates required by the query.

However, there is no discussion of vertically overlapping automatic summary tables in Chiang. Indeed, Chiang is directed only to the construction of a compressed histogram of a subject table without reference to a query.

Consequently, Chiang does not teach or suggest "generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query."

The Office Action also states that Schiefer teaches the elements "using the cardinality estimates to determine an optimal query execution plan for the query" at col. 3, lines 37-60 and col. 10, lines 23-57. However, at the indicated locations, Schiefer merely states the following:

> <u>Schiefer: Col. 3, lines 37-60</u>
>
> It is another object of the present invention to produce a better cardinality estimate by utilizing information and attributes which can be obtained from the catalog for the relational database management system. The additional information includes cardinalities for existing unique keys, column equivalence classes, functional dependencies, statistical functional dependencies, and statistically unique keys.
>
> In a first aspect, the present invention provides a method for estimating cardinalities for a key formed from a grouping of columns in a table for use in a query optimizer for a relational database management system, wherein selectivities and keys associated with columns in the table are provided in a catalog, said method comprising the steps of: (a) determining an equivalence class for each column in said key; (b) for each said equivalence class determining an effective cardinality for each of said columns belonging to said equivalence class; (c) determining a cardinality for each of said equivalence classes by choosing the minimum effective cardinality for the columns belonging to said equivalence class; and (d) estimating a cardinality value for said key from the product of said cardinalities for said equivalence classes.
>
> <u>Schiefer: Col. 10, lines 23-57</u>
>
> To determine the effective cardinality of a column in Line 13, the method according to the present invention considers the effect of local predicates on other columns in the equivalence class. Known query optimizers estimate the cardinality of a column C1 using only the product of predicate selectivity (ff_1) and base table column cardinality .vertline.C1.vertline. obtained from the CATALOG. Known optimizers do not consider the effects of predicates on other columns. According to the invention, the effective cardinality of a column is determined by the following expression which will be referred to as Expression (1):

EFFECTIVE COLUMN CARDINALITY = $|C1| * ff\_1 * (1 - (1 - ff\_2)^{(|T|/|C1|)})$(1)

where:

$|T|$ is the table cardinality, i.e. number of rows in table

$|C1|$ is the base table cardinality obtained from the CATALOG

ff_1 is the selectivity of a local predicate for column C1

ff_2 is the selectively of a local predicate for column C2

In the derivation of Expression (1) according to the present method, it is assumed that C1 and C2 are independent, and the values of C1 and C2 are both uniformly distributed in the table.

If there is no restriction on column C2, i.e. ff_2 is 1, Expression (1) reduces to $|C1| * ff\_1$ which provides the basic operation performed by known optimizers for obtaining the effective cardinality of a column. Since the prior art method is based on the assumption that all columns in a key are fully independent of each other, the method according to the prior art usually leads to unnecessarily large numbers for the key cardinalities. This in turn can result in the query optimizer 18 (FIG. 1) picking the wrong query plan which is clearly undesirable.

In the context of Appellant's claims, the cardinality estimates are generated using statistics of one or more automatic summary tables that vertically overlap the query. In Schiefer, however, the cardinality estimates are generated by (1) determining an equivalence class for each column in a key; (b) for each equivalence class, determining an effective cardinality for each of the columns belonging to the equivalence class; (c) determining a cardinality for each of the equivalence classes by choosing the minimum effective cardinality for the columns belonging to the equivalence class; and (d) estimating a cardinality value for the key from the product of the cardinalities for the equivalence classes.

Raitto does not overcome the deficiencies of Schiefer and Chiang. Recall that Raitto was cited only against dependent claims 6-10, 16-20 and 26-30, and is specifically directed to queries that do not reference a particular materialized view (automatic summary table).

Consequently, even when combined, the Schiefer, Chiang and Raitto references do not teach or suggest the Appellant's invention. Moreover, the various elements of Appellant's claimed invention together provide operational advantages over the cited references. In addition, Appellant's invention solves problems not recognized by the cited references.

Thus, Appellant submits that independent claims 1, 11 and 21 are allowable over Schiefer, Chiang and Raitto. Further, dependent claims 3-10, 13-20 and 23-30 are submitted to be allowable

over Schiefer, Chiang and Raitto in the same manner, because they are dependent on independent claims 1, 11 and 21, respectively, and because they contain all the limitations of the independent claims.

G. Appellant's Dependent Claims Are Patentable Over The Cited References

Appellant's dependent claims are patentable over Schiefer, Chiang and Raitto, because it includes a combination of limitations not taught or suggested by the cited references, taken individually or in any combination

With regard to claims 3, 13 and 23, which recite that "the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query," the Office Action asserts that Schiefer teaches these elements at col. 6, line 41 – col. 7, line 20. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes estimating cardinalities, but not using statistics of automatic summary tables.

With regard to claims 4, 14 and 24, which recite that "the predicates are applied by one of the automatic summary tables," the Office Action asserts that Schiefer teaches these elements at col. 10, lines 23-44. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes the effect of local predicates on other columns, but not the application of predicates by automatic summary tables.

With regard to claims 5, 15 and 25, which recite that "the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query," the Office Action asserts that Schiefer teaches these elements at col. 8, lines 1-28. Appellant's attorney disagrees. At the indicated location, Schiefer merely describes estimates of key cardinalities, but says nothing about a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

With regard to claims 6, 16 and 24, which recite that "zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table," the Office Action asserts that Chiang teaches these elements at col. 12, lines 28-67. Appellant's attorney disagrees. The indicated location in Chiang is completely unrelated to the claim limitations.

With regard to claims 7, 17 and 27, these claims stand or fall with claims 1, 11 and 21.

With regard to claims 8, 18 and 28, which recite "determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table," the Office Action asserts that Raitto teaches these elements at col. 11, lines 6-19. Appellant's attorney disagrees. The indicated location in Raitto describes a "query reduction factor" computed for a materialized view, which is a different ratio, comprising a ratio of (1) the sum of the cardinalities of matching relations in the query that will be replaced by the materialized view to (2) the cardinality of the materialized view.

With regard to claims 9, 19 and 29, which recite that "a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query," the Office Action asserts that Raitto teaches these elements at col. 11, lines 20-30. Appellant's attorney disagrees. At the indicated location, Raitto describes a "query reduction factor" computed for a materialized view, which is a different ratio, comprising a ratio of (1) the sum of the cardinalities of matching relations in the query that will be replaced by the materialized view to (2) the cardinality of the materialized view.

With regard to claims 10, 20 and 30, which recite that "the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio," the Office Action asserts that Raitto teaches these elements at col. 11, lines 31-49. Appellant's attorney disagrees. At the indicated location, Raitto merely describes the "current best materialized view" with the "highest query reduction factor," but not a product of the subpredicate combined selectivity estimate and the cardinality ratio.

IX. CONCLUSION

In light of the above arguments, Appellant respectfully submits that the cited references do not anticipate nor render obvious the claimed invention. More specifically, Appellant's claims recite novel physical features, which patentably distinguish over any and all references under 35 U.S.C. §§ 102 and 103. As a result, a decision by the Board of Patent Appeals and Interferences reversing the Examiner and directing allowance of the pending claims in the subject application is respectfully solicited.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Appellant

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: May 21, 2004

By: ____________________
Name: George H. Gates
Reg. No.: 33,500

GHG/

G&C 30571.272-US-U1

APPENDIX

1. A method of optimizing execution of a query that accesses data stored on a data store connected to a computer, comprising:

generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query; and

using the generated cardinality estimates to determine an optimal query execution plan for the query.

3. The method of claim 1, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

4. The method of claim 3, wherein the predicates are applied by one of the automatic summary tables.

5. The method of claim 4, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

6. The method of claim 3, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

7. The method of claim 6, wherein a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

8. The method of claim 7, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

9. The method of claim 8, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

10. The method of claim 9, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

11. An apparatus for optimizing execution of a query, comprising:

a computer having a data store coupled thereto, wherein the data store stores data;

one or more computer programs, performed by the computer, for generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query, and for using the generated cardinality estimates to determine an optimal query execution plan for the query.

13. The apparatus of claim 11, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

14. The apparatus of claim 13, wherein the predicates are applied by one of the automatic summary tables.

15. The apparatus of claim 14, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

16. The apparatus of claim 13, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

17. The apparatus of claim 16, a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

18. The apparatus of claim 17, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

19. The apparatus of claim 18, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

20. The apparatus of claim 19, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.

21. An article of manufacture comprising a program storage medium readable by a computer and embodying one or more instructions executable by the computer to optimizing execution of a query that accesses data stored on a data store connected to the computer, comprising:

generating cardinality estimates for one or more query execution plans for the query using statistics of one or more automatic summary tables that vertically overlap the query; and

using the generated cardinality estimates to determine an optimal query execution plan for the query.

23. The article of manufacture of claim 21, wherein the statistics of the one or more automatic summary tables are used to improve a combined selectivity estimate of one or more predicates of the query.

24. The article of manufacture of claim 23, wherein the predicates are applied by one of the automatic summary tables.

25. The article of manufacture of claim 24, wherein the selectivity estimate comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

26. The article of manufacture of claim 23, wherein zero or more predicates of the query are applied by one of the automatic summary tables and wherein the remaining predicates are eligible to be applied on the automatic summary table.

27. The article of manufacture of claim 26, a predicate is eligible to be applied on the automatic summary table if it can be evaluated using the output columns and expressions of the automatic summary table.

28. The article of manufacture of claim 27, further comprising determining a subpredicate combined selectivity estimate of the unapplied eligible predicates using column distribution statistics of the automatic summary table.

29. The article of manufacture of claim 28, wherein a cardinality ratio comprises a ratio of a cardinality of the automatic summary table to a product of cardinalities of base tables referenced in the automatic summary table and the query.

30. The article of manufacture of claim 29, wherein the selectivity estimate comprises a product of the subpredicate combined selectivity estimate and the cardinality ratio.